Applicability and clinical relevance of results in randomized controlled trials : the Cochrane review on exercise therapy for low back pain as an example

Malmivaara, A.; Koes, R.E.; Bouter, L.M.; van Tulder, M.

published in
Spine
2006

DOI (link to publisher)
10.1097/01.brs.0000219868.30427.66

document version
Publisher's PDF, also known as Version of record

SPINE Volume 31, Number 13, pp 1405–1409

Applicability and Clinical Relevance of Results in Randomized Controlled Trials

The Cochrane Review on Exercise Therapy for Low Back Pain as an Example

Antti Malmivaara, MC, PhD,* Bart W. Koes, PhD,† Lex M. Bouter, PhD,‡ and Maurits W. van Tulder, PhD‡

Study Design. A critical appraisal of the literature.

Objectives. To increase awareness of the importance of applicability and clinical relevance of the results of randomized controlled trials (RCTs) in the field of spinal disorders by formulating a list of items for assessment of applicability and clinical relevance of results of RCTs.

Summary of Background Data. In systematic reviews of randomized controlled trials (RCTs), critical appraisal of methodologic quality is considered important. Less attention has been paid to the assessment of the applicability and the clinical relevance of the results.

Methods. RCTs in an update of the Cochrane review on exercise therapy for low back pain were used. Most of the trials did not score positively on the five Cochrane Back Review Group basic items describing patients: intervention and setting, outcome, effect size, and benefits related to adverse effects. Item 1 was met by 88% of the trials, but item 2 only by 51%, item 3 by 67%, item 4 by 35%, and item 5 by 0%. Subsequently, a more comprehensive list of items for the assessment of applicability and clinical relevance of results of RCTs was developed. These criteria were pilot tested on the RCTs. After pilot testing and a subsequent consensus meeting, the list of items was drafted and circulated among the members of the Editorial Board of the Cochrane Back Review Group. Changes were made in response to comments.

Results. The final list consists of 40 items. The items are ordered on two headings: Does the report enable the assessment of applicability? Are the study results clinically relevant? We present examples of informative and noninformative reporting of RCTs in order to illustrate how information on applicability and clinical relevance of results can be assessed.

Conclusions. Authors of RCTs should adequately report on items that are essential to assess the applicability and clinical relevance of results. The presented list of items may help clinicians reading RCTs and authors of systematic reviews to draw more balanced conclusions on applicability and clinical relevance of results.

Key words: randomized controlled trial, back pain, applicability, clinical relevance, methodology. **Spine 2006; 31:1405–1409**

Randomized controlled trials (RCTs on treatment efficacy) constitute the backbone of evidence-based medicine. Evidence-based medicine involves conscientious, explicit, and judicious use of current best evidence in making decisions about care of individual patients.[1] Five steps have been suggested for using evidence-based medicine in clinical practice: 1) ask clinical questions you can answer, 2) search for the best evidence, 3) critically appraise the evidence, 4) apply the evidence in care for your patient, and 5) self-evaluation of the above steps. Searching for the best literature and critically appraising the evidence are essential in conducting systematic reviews. For clinicians in daily practice, it is impossible to systematically identify, critically appraise, and summarize the literature because of the tremendous number of scientific papers published each year. Systematic reviews offer clinicians a solution to this problem. A lot of attention has been paid to improving the methods of systematic reviews, making them more valid and reproducible. However, when clinicians need to make the decision if the evidence from RCTs or systematic reviews of RCTs can be applied to their patients also, the applicability and clinical relevance of the results are important.

Applicability and clinical relevance of the results deal with the question of whether and how to use the evidence in practice. Both aspects deserve serious attention in systematic reviews. However, the design, conduct, and report of trials have not been optimal. Since publication of the CONSORT statement, most major clinical scientific journals, including *Spine*, have adopted these recommendations to improve the quality of reporting of RCTs.[2] Attention has been paid to reporting of RCTs to enable and improve the critical appraisal of the quality of RCTs in terms of their internal validity, but to a much lesser extent to the assessment of their applicability and the clinical relevance of the results. Furthermore, clinicians sometimes suggest that RCT results are not clinically relevant. In a recent update of their method guidelines, the Cochrane Back Review Group recommended includ-

From the *Finnish Office for Health Technology Assessment, FinOHTA/Stakes, Helsinki, Finland; †Department of General Practice, Erasmus Medical Center, Rotterdam, the Netherlands; and ‡Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, the Netherlands. Members of the Editorial Board of the Cochrane Back Review Group, who contributed to this paper, are as follows: Lex M. Bouter, PhD; P. G. Shekelle, MD, MPH; R. A. de Bie, PhD; F. Guillemin, MD, PhD; M. Schoene, MD, PhD; and Maurits W. van Tulder, PhD.

Acknowledgment date: August 27, 2004. First revision date: April 15, 2005. Second revision date: July 4, 2005. Third revision date: August 10, 2005. Acceptance date: August 15, 2005.

The manuscript submitted does not contain information about medical device(s)/drug(s).

Other funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Antti Malmivaara, MD, PhD, Finnish Office for Health Technology Assessment, Helsinki, Finland; E-mail: antti.malmivaara@stakes.fi

ing an assessment of clinical relevance of study results in systematic reviews in the field of spinal disorders. Five questions that are based on previous guides to critical reading of the medical literature were recommended[3]:

1. Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?
2. Are the interventions and treatment settings described well enough so that you can provide the same for your patients?
3. Were all clinically relevant outcomes measured and reported?
4. Is the size of the effect clinically important?
5. Are the likely treatment benefits worth the potential adverse effects?

These five questions assess two dimensions:

- Does the report enable assessment of applicability? This is reflected in the questions 1 to 3.
- Are the study results clinically relevant? This is the main focus of questions 4 and 5.

Both dimensions can only be assessed if the relevant information is clearly reported in publications of RCTs. Also, when questions 1 to 3 are answered positively, this does not necessarily mean that the RCT is clinically relevant. The report can, for instance, make it very clear that the patients included in the RCT are very different from the patients the clinician sees in his own practice.

We used the five questions to assess applicability and clinical relevance of the results of RCTs included in an updated Cochrane review of exercise therapy for low back pain.[4] Most of the trials did not score positively on these items. Item 1 was met by 88% of the trials, but item 2 only by 51%, item 3 by 67%, item 4 by 35%, and item 5 by 0%. We thus found it difficult to assess and judge the clinical relevance of the RCTs and decided to operationalize the five questions further.

We assessed how the applicability and clinical relevance of the results were reported in the update of the systematic Cochrane review piloting a more comprehensive list of relevant items. The results were circulated among members of the Cochrane Back Review Group editorial board members and in response to the comments a final list of items was constructed.

The aim of the present paper is to increase awareness of the importance of applicability and clinical relevance of the results of RCTs in the field of spinal disorders and to improve the reporting RCTs and their results in field of spinal research.

The final consensus consisted of 40 items that are related to applicability of the trial and clinical relevance of the results (Table 1). These final items are explained below with examples of informative and noninformative reporting in trials included in the update of the Cochrane systematic review on exercise therapy for low back pain.[4]

We have picked the examples for illustration purposes only. The citations have been taken out of the context and do not by any means disqualify the RCTs from which they come.

Table 1. Reporting of Items Related to Applicability and Clinical Relevance of Results of RCTs

Methods: Does the report enable the assessment of applicability?
Study population
1. Age
2. Gender
3. Setting
4. Type of disease/disorder
5. Duration of disease/disorder
6. Severity of disease/disorder
7. Recruitment procedure
8. Description of inclusion and exclusion criteria
Index intervention
9. Type/content
10. Intensity/dosage
11. Frequency
12. Duration
13. Experience of provider
14. Proper intervention to answer the research question
Comparator (control intervention)
15. Type/content
16. Intensity/dosage
17. Frequency
18. Duration
19. Experience of provider
20. Proper control group to answer the research question
Cointerventions per study group
21. Type/content
22. Intensity/dosage
23. Frequency
24. Duration
25. Experience of provider
Outcome measures
26. Main symptom, disease-specific disability, and generic disability
27. Validity and reliability of instruments
28. Follow-up moment
29. All potential adverse effects
Analysis
30. Intention-to-treat analysis
31. Confounding considered
32. Effect modification considered
33. Economic evaluation
Results: Are the study results clinically relevant?
34. Baseline values of main symptoms and disability plus measure of variance
35. Adherence in all study groups
36. Dropout rate
37. Follow-up values of main symptoms and disability plus measure of variance
38. Confidence intervals of between-group differences
39. Magnitude of difference between groups
40. Incidence of all adverse effects

■ Methods

Does the report enable the assessment of applicability?

Study Population. An explicit description of the age and gender (or any other relevant demographic data) of the study population and the type, duration, and severity of the disease or disorder will enable clinicians to decide if the results of this trial also relate to their patient population. The setting in which the trial was conducted should be clearly described as well as the recruitment procedure. Patients recruited from the records of a general practitioner are most likely different from patients recruited in the waiting room of an orthopedic surgeon or patients recruited through advertisements in a newspaper. Ex-

plicit description of inclusion and exclusion criteria gives the reader information on the specific population that was included in the trial and consequently enables comparison with their own patients. Sometimes these differences are relevant, sometimes not. RCTs results can be valid for patients who would not be eligible for the RCT at issue. The central question is whether the differences are effect modifying characteristics.

In case of pragmatic trials, it should be clear if patients with (strong) treatment preference are included or excluded from the trial. If patients with (strong) treatment preferences are included, this may be a confounding factor which should be dealt with (item 32).

Example of Informative Reporting. "Patients who consulted their general practitioners for back pain were selected. Inclusion criteria were pain between T12 and the gluteal fold with or without radiation to the upper leg, pain for 3 weeks or less, and age between 16 and 65 years. Exclusion criteria were . . . " The authors also included details of baseline characteristics of demographics and duration and severity of back pain for people in all study groups in a table.

Example of Noninformative Reporting. "Patients had to be referred by a physical medicine or orthopaedic specialist. Only those agreeing to attend initially on a regular basis for 8 to 14 treatment sessions and to be available for subsequent follow-up assessments at a 3-month then a 12-month interval were included for the survey." The author also reported a patient profile with some baseline characteristics of the entire study population, but not per study group.

Index Intervention. To be able to reproduce an intervention and to apply the intervention to your own patients, the type and content of the intervention, the intensity or dosage, and the frequency and duration should be explicitly described. A description of the provider of the intervention and, if relevant, his or her training and experience is needed for clinicians to judge if they would be able to provide this intervention themselves or if additional training is needed. For this the authors may refer to another source outside the article at issue that provides full information on the intervention. The intervention may not be necessary standardized according to a predefined protocol; *e.g.*, in case of folk and traditional therapy, there may be wide variation in the intensity, type, frequency, and duration of the therapy. However, interventions that vary from patient to patient should be described adequately so that the intervention is reproducible by others in clinical practice. In these cases, treatment algorithm or decision models should be presented.

Example of Informative Reporting. "Before commencing the exercise programs, patients were shown how to use the training programs and supervised carefully by the physiotherapist in charge . . . The two exercise programs are shown in Figures 2 and 3. Each of the training programs consisted of 9 basic exercises that were to be carried out in 3 series of 10 repetitions three times a week. Progression in the programs, decided in cooperation with the physiotherapist, was done by adding extra weights when carrying out the exercises." Examples of the types of exercises are presented in these figures.

Example of Noninformative Reporting. "The physiotherapy consisted of a combination of manual, thermal, and electrotherapy. The therapist was free to choose a suitable method within these categories and to use the facilities at his disposal: hot/cold packs, infrared heat, ultrasound, short-wave diathermy, and transcutaneous electric nerve stimulation. In addition to massage, he also employed specific mobilizations and manual traction according to the GPs prescription, but no manipulations with impulse. Individual autostretching exercises were added if tightness of the pelvic or femoral muscles was noted at the initial examination."

Control Intervention. Different control interventions give answers to different research questions. Also for the control intervention, it is important to know what the type and content, intensity or dosage, frequency, and duration were and if providers needed experience or special training.

Example of Informative Reporting. "Placebo therapy: ultrasound of 20 minutes' duration, at the lowest possible dose (0.1 W/cm^2, intermittent) was administered by a physiotherapist twice weekly for 5 weeks. Because of the intermittent character and the minimal level of the dose without heat effect, this treatment was considered a placebo treatment. A record of each treatment was kept by a physiotherapist."

Example of Noninformative Reporting. "Short-wave diathermy (SWD) to the lumbosacral spine . . . Treatments were continued for a 4-week period under the supervision of the same physiotherapist."

Cointerventions. Even if the intervention that is evaluated in a trial is clearly described, one needs to know if there were any cointerventions during the intervention period. If patients also received other interventions, information on the type and content, intensity or dosage, and frequency and duration of these other interventions should also be provided.

Example of Informative Reporting. "Eighteen percent of the subjects in the booklet group visited a health care provider for back pain during the study month, and only 8% of the subjects in the chiropractic group and 9% of those in the physical therapy group visited providers other than those assigned. The reported use of exercise was almost identical in the three groups at baseline (57%) and 1 month (about 81%). During the month, the percentage of subjects who used back-pain medication of any type decreased from 82% to 18% in the chiropractic group, from 84% to 27% in the physical therapy group, and from 77% to 32% in the booklet group ($P < 0.05$ for the differences among the groups after adjustment for baseline use). Fewer than 2% of the subjects reported using corsets, braces, traction, TENS, or injections."

Example of Noninformative Reporting. It is difficult to present an example of noninformative reporting because the majority of reports of RCTs (84%) evaluating exercise therapy for low back pain do not present any information on cointerventions at all (neither in the Methods nor in the Results section). Some studies suggest in their methods that patients were allowed to continue using their regular pain medication but did not report the actual use of medication in the results. Also, medication is only one of the possible cointerventions and other healthcare utilization during the intervention period should be measured and reported as well.

Outcome Measures. A core set of primary outcome measures in the field of low back pain has been proposed.[2] Trials that include main symptom, disease-specific disability, and generic disability are more relevant than trials that do not include these

Table 2. Example of Informative Reporting of Results: Group Differences in Primary Outcome Measures After 11 Wk of Intervention/Exercise (Means are Adjusted for Baseline Values for Comparability)

	Group A SMT + Strength	Group B NSAID + Strength	Group C SMT + Stretch	Group Differences (95% CI) A–B	Group Differences (95% CI) A–C
Pain: 11-box scale (SD)	2.7 (2.0)	3.5 (2.2)	3.3 (2.3)	0.8 (−0.02; 1.6)	0.6 (−0.2; 1.4)
Disability: Roland-Morris (SD)	15.1 (17.4)	20.9 (17)	18.4 (17.1)	5.8 (−1.1; 12.7)	3.3 (−3.6; 10.2)
General health: COOP (SD)	75.4 (12)	75.6 (11.1)	74.8 (16.3)	0.2 (−4.2; 4.6)	−0.6 (−4.8; 4.0)

measures. One could even argue that trials that do not include any outcomes of symptom and disability are clinically irrelevant. Outcomes should be measured with valid, reliable and responsive instruments.

Follow-up moments should be frequent and the duration of follow-up long enough to monitor the expected effect. Trials on interventions for chronic low back pain that aim to reduce work absenteeism should have at least 1-year follow-up.

All potential adverse effects should be monitored.

Example of informative reporting. "Measures of outcome were grouped in four categories: functional status, which included a modified Sickness Impact Profile (a comprehensive health-status questionnaire previously validated for use in low back pain) . . ., pain ratings . . ., physical measures . . ., use of medical services . . ."

Example of Noninformative Reporting. "Before the intervention, all patients rated back pain on a visual analog scale (VAS) as applied by Huskissen ranging from 0 (no pain) to 10 (severe pain). Resting pulse, blood pressure in lying position, height, body weight, socioeconomic status, underlying disease, level of education, occupation, activity and serum for high density lipoprotein-cholesterol (HDL-C) were also recorded. The scoring procedure was repeated at the final follow-up examination (after 12 weeks of intervention)."

Analysis. If data in trials are analyzed according to the intention-to-treat principle, they are more relevant. Potential and actual confounding as well as effect modification should be considered. In case of pragmatic trials, inclusion of patients with treatment preferences may introduce a bias. Economic evaluations are becoming more and more popular. Knowing if an intervention is not only effective but also what the incremental costs of this intervention are, facilitates the decision to use this intervention in daily practice or not. Therefore, adequate reporting of economic evaluations is important.

Example of Informative Reporting. "Our analysis was based on intention to treat. We estimated the effects of treatment on the outcome measures by means of analysis of covariance, with the change in scores as the dependent variable and adjustment being made for baseline score and patient preference. We used Student's *t* test to analyze the data from pain diaries as the baseline scores were quite similar."

" . . . Economic analysis. We recorded patients' use of healthcare services using a combination of retrospective questionnaires and prospective diary cards, which they returned at 6 and 12 months' follow up. From this information we estimated the cost of each patients' treatment. We compared the mean costs of treatment for the two groups by using Student's *t* test and standard confidence intervals. However, as cost data were highly positively skewed, these results were checked with a nonparametric 'bootstrap.' The economic evaluation addressed both costs to NHS and the costs to society. Participants were not charged for the classes, in line with any treatment currently available on the NHS."

Example of Noninformative Reporting. "Statistical analysis. The χ^2 test for two independent samples, Fisher's exact test, the Mann-Whitney U test, and the Wilcoxon signed rank test were used as appropriate for statistical analysis of the data, with P values <0.05 considered statistically significant."

■ Results

Are the study results clinically relevant?
Interpretation of the results of a trial is only possible if baseline values of the main symptoms and disability are presented, including a measure of variance. If baseline values are unknown, it is unclear for which patients the intervention may be useful. Adherence in all study groups should be reported. An intervention with a low adherence rate should not be directly implemented. The magnitude of the difference between groups and confidence intervals of this between-group difference during follow-up should be reported. All adverse effects should be reported.

Example of Informative Reporting

Group differences in primary outcome measures after 11 weeks of intervention/exercise (means are adjusted for baseline values for comparability) (Table 2).

The authors also presented baseline data (mean and standard deviations) of outcome measures in another table as well as results on compliance, cointerventions, and dropout rates.

Example of Noninformative Reporting

"*All patients*. Throughout the observation period, a significant ($P < 0.01$) reduction of pain was registered within all three total-sample treatment groups. (Pain level and overall treatment effect were calculated over time, and only patients attending all evaluations were included.) There were no significant differences in pain level among the three total-sample treatment groups at any time." The authors did not report intention-to-treat data but only presented data of subgroups.

■ Discussion

The clinical implications of an RCT can only be assessed if authors clearly describe the items enabling the reader to interpret whether the results are applicable to a par-

ticular clinical situation, and whether the results make treatment worth of use in clinical practice. The examples of the systematic review of exercise therapy for low back pain show that many publications still lack relevant information and that there is a dire need for improving the quality of reporting to enable the assessment of applicability of the RCT and clinical relevance of its results. This problem has also been confirmed in the field of cardiovascular disorders.[5]

The CONSORT statement included recommendations for reporting of RCTs to enable readers to differentiate trials with unbiased results from those that are biased. Assessment of internal validity is very important for the interpretation of trials. However, another important aspect that is often underestimated is applicability and clinical relevance of the results (do the results of the trial make a difference for clinical practice?). A high-quality trial may be irrelevant to clinicians. For example, a trial may show validly that exercise therapy is more effective than spinal manipulation for improving range of motion in chronic patients with low back pain. Only if clinically relevant outcomes of symptoms and/or disability are also measured will this trial have impact on clinical management. Ultimately, reports of trials should combine information necessary to assess the internal validity of the trials with information necessary to assess its applicability and clinical relevance of results.

Whether a specific RCT is applicable and its results are clinically relevant may depend on the particular circumstances, as well as on the values and preferences of the particular patient. Thus, the main difference between internal validity on one hand, and applicability and clinical relevance, on the other hand, is that the former deals with evidence that is in its essence abstract and has a potential for wide applicability over place and time. The latter is a prerequisite for the abstract knowledge to be applied into a particular clinical situation, *i.e.*, an informative reporting of applicability and clinical relevance makes it possible to judge whether application of the trial evidence is worthwhile in a specific situation. A trial may be relevant to an orthopedic surgeon but not for a physiotherapist, or a trial may be relevant for a policy maker but not for a clinician and his or her patient.

We think that it is essential for readers to be able to assess the most important items related to the applicability and clinical relevance of results of RCTs. This paper presents a list of items that may help improve the reporting of RCTs, and may also be helpful for clinicians reading reports of RCTs.

■ Key Points

- Randomized controlled trials (RCTs) should adequately report on items that are essential to assess the applicability (generalizability) and clinical relevance of results.
- The presented list of items may help clinicians reading RCTs and authors of systematic reviews to draw more balanced conclusions on applicability and clinical relevance of results.

References

1. Sackett D. *Evidence Based Medicine*. London: Churchill Livingstone, 1997.
2. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
3. Shekelle PG, Andersson G, Bombardier C, et al. A brief introduction to the critical reading of the clinical literature. *Spine* 1994;19(suppl):2028–31.
4. Hayden JA, van Tulder MW, Malmivaara AV, et al. Meta-analysis: exercise therapy for nonspecific low back pain. *Ann Intern Med* 2005;142:765–75.
5. Willenheimer R, Dahlöf B, Gordon A. Clinical trials in cardiovascular medicine: are we looking for statistical significance or clinical relevance? *Heart* 2000;84:129–33.